Supplementary Information

A widespread self-cleaving ribozyme class is revealed by bioinformatics

Adam Roth[1,2*], Zasha Weinberg[1,2*], Andy G. Y. Chen[1] , Peter B. Kim[1] , Tyler D. Ames[1,2,†] & Ronald R. Breaker[1,2,3]

[1] Department of Molecular, Cellular and Developmental Biology, [2]Howard Hughes Medical Institute, [3]Department of Molecular Biophysics and Biochemistry, Yale University, Box 208103, New Haven, CT 06520-8103, USA.

* These authors contributed equally to this work.

[†]Current Address: Phosplatin Therapeutics, 1350 Avenue of the Americas, New York, NY 10019, USA.

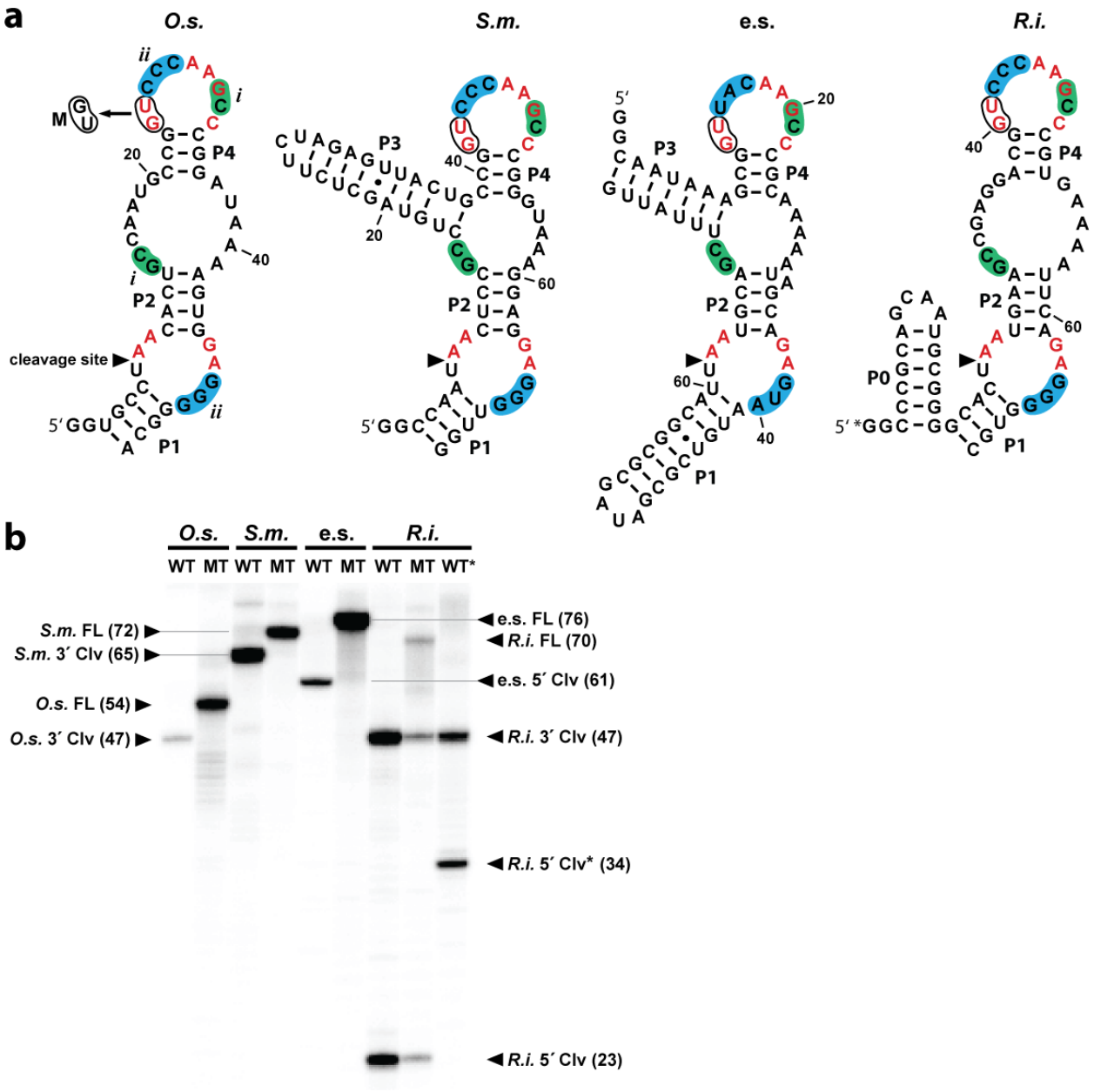
Supplementary Results

Supplementary Data Set 1

This separate file displays the taxa, sequence accession numbers, coordinates, gene neighborhoods and multiple sequence alignments corresponding to the identified twister ribozyme representatives. In additional separate files, alignments of type P1, type P3 and type P5 twister ribozyme sequences are provided individually in Stockholm formats.

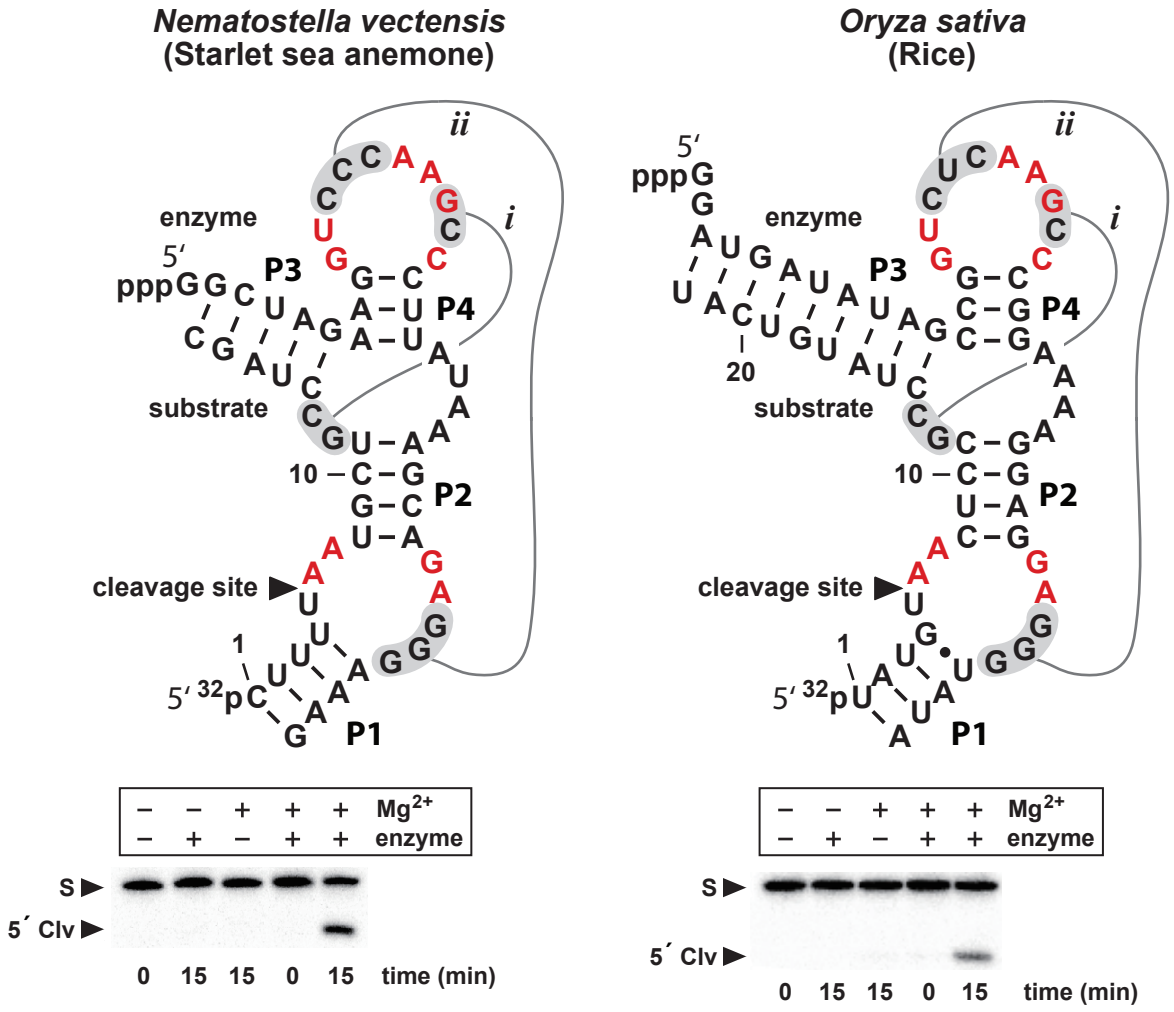
Supplementary Table 1 | Genomic distributions of twister ribozymes. Organisms known to contain predicted type P1 twister ribozymes are listed with the corresponding numbers of examples. Note that only convincing ribozyme homologs are included, as judged by the degree to which they conform to the conserved sequence and structural features depicted in **Fig. 1a**. Additional twister representatives with sequence or structural variations could exist, and therefore the numbers provided in this table might be underestimates. Not included in this table are 1,495 twister ribozymes derived from metagenome DNA sequences (1,155 type-P1, nine type-P3 and 331 type-P5 twister ribozymes), and seven type P5 twister ribozymes that were identified in species of Clostridia or in *Strongylocentrotus purpuratus*.

Domain	Lineage	Organism	Number of twister ribozymes
Bacteria	Clostridia	*Blautia hydrogenotrophica* DSM 10507	1
		Clostridium bolteae ATCC BAA-613	3
		Clostridium botulinum A str. ATCC 3502	1
		Clostridium botulinum A2 str. Kyoto	1
		Clostridium botulinum Ba4 str. 657	1
		Clostridiales bacterium 1 7 47FAA	2
		Roseburia intestinalis L1-82	1
		Anaerotruncus colihominis DSM 17241	1
		Ethanoligenens harbinense YUAN-3	1
		Ruminococcus gnavus ATCC 29149	2
		Subdoligranulum variabile DSM 15176	2
		Ruminococcaceae bacterium D16	2
	Planctomycetacia	*Gemmata obscuriglobus* UQM 2246	8
	(phage)	*Clostridium* phage c-st	1
Eukarya	fungus	*Phaeosphaeria nodorum* SN15	1
	insect	*Nasonia vitripennis*	47
		Acyrthosiphon pisum	17
	fish	*Danio rerio*	10
	cnidarian	*Nematostella vectensis*	2
		Hydra magnipapillata	1
	flatworm	*Schistosoma mansoni*	1,051
	plant	*Ricinus communis*	4
		Oryza sativa Japonica Group	8
		Populus trichocarpa	5
		Sorghum bicolor	15

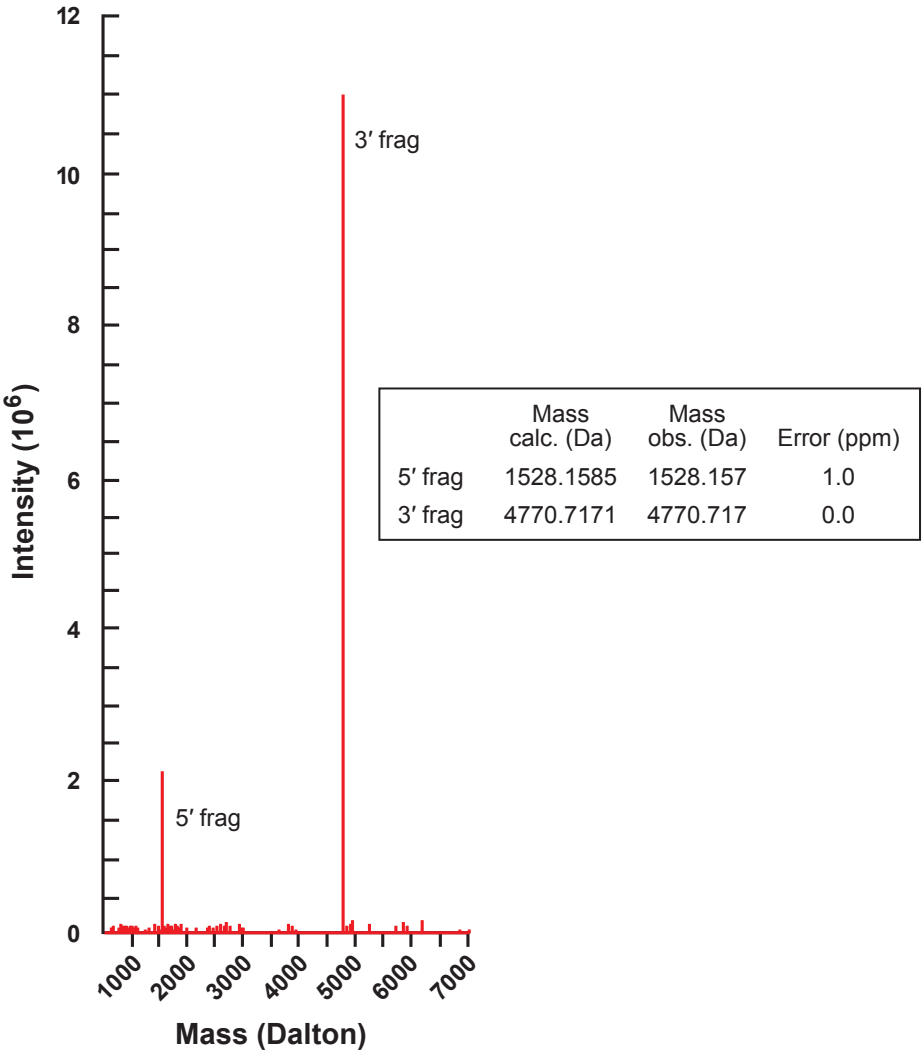


Supplementary Figure 1 | Twister ribozymes from different kingdoms self-cleave during transcription *in vitro*. (a) Twister RNAs from *Oryza sativa* (*O.s.*), *Schistosoma mansoni* (*S.m.*), environmental sequence (e.s.), and *Roseburia intestinalis* (*R.i.*) were used to assess self-cleavage activity during transcriptions *in vitro*. Pseudoknot-forming sequences *i* and *ii* are highlighted in green and blue, respectively. For each construct, presumed disruptive mutations were engineered by altering the same conserved nucleotides (circled), as shown for *O.s.* The asterisk refers to a

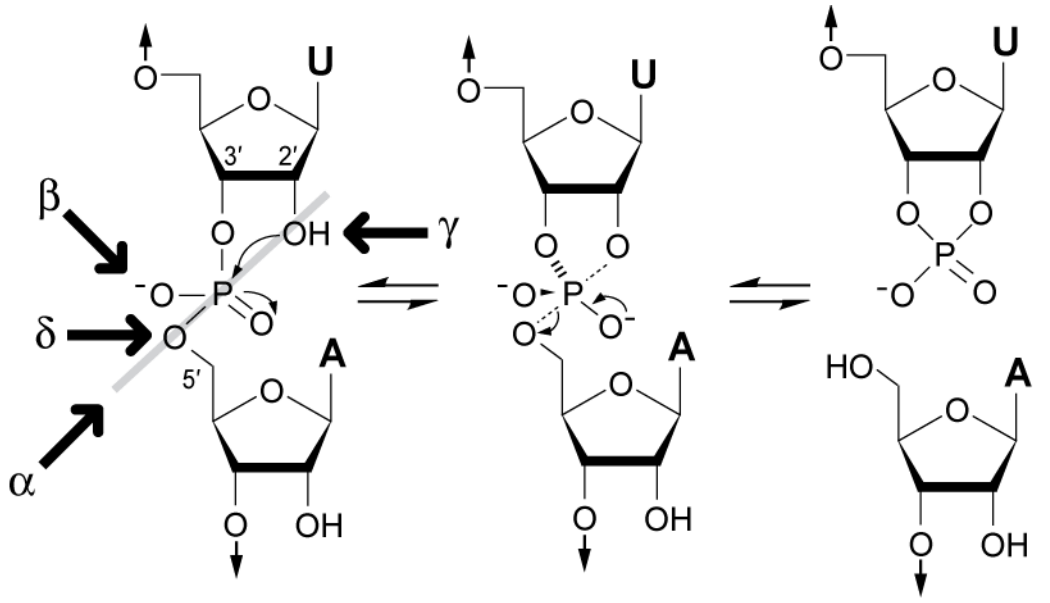
second wild-type *R.i.* construct (not shown) that contains an additional 11 nucleotides at the 5′ end. Non-native guanosine residues were added to the 5′ ends to facilitate transcription *in vitro*. (b) The internally ^{32}P-labeled products (nucleotide lengths in parentheses) generated during *in vitro* transcriptions of wild-type (WT) and mutant (MT) versions of the constructs in (a) were separated by denaturing 15% PAGE. The full length transcripts (FL) and 5′ and 3′ cleavage products (5′ Clv and 3′ Clv, respectively) are indicated. Note that the 5′ products generated by the *O.s.* and *S.m.* constructs and the 3′ product generated by the e.s. construct were obscured due to their approximate co-migration with unincorporated α-^{32}P [GTP]; this part of the image is not shown. In contrast, the longer 5′ cleavage products generated by the *R.i.* constructs are readily detected. The residual activity of the mutant *R.i.* construct might result in part from structural stabilization of the disrupted ribozyme core by the P0 stem.



Supplementary Figure 2 | Bimolecular constructs (top) and ribozyme cleavage assays (bottom) for two twister representatives. Nucleotides in red are conserved in at least 97% of identified twister ribozymes. Non-native guanosine residues were added to the 5′ ends of the enzyme strands to facilitate transcription *in vitro*. Reaction mixtures contained the appropriate RNAs, 20 mM HEPES (pH 7.5 at 23°C), 100 mM NaCl, and 20 mM $MgCl_2$, except in cases where $MgCl_2$ or the enzyme strand was omitted. 5′ ^{32}P-labeled substrates (S) were separated from the corresponding 5′ cleavage fragments (5′ Clv) by denaturing 20% PAGE.



Supplementary Figure 3 | Mass spectrum analysis of the products of a bimolecular twister ribozyme cleavage reaction. The enzyme and substrate RNAs, which are derived from *C. bolteae*, are depicted in **Supplementary Fig. 8a**. The reaction product masses are consistent with the cleavage site and end groups indicated in **Fig. 3a** and **Fig 3c**, respectively. The error, in parts per million (rounded to 0.1), is the difference between the calculated and observed atomic masses.



Catalytic Strategy

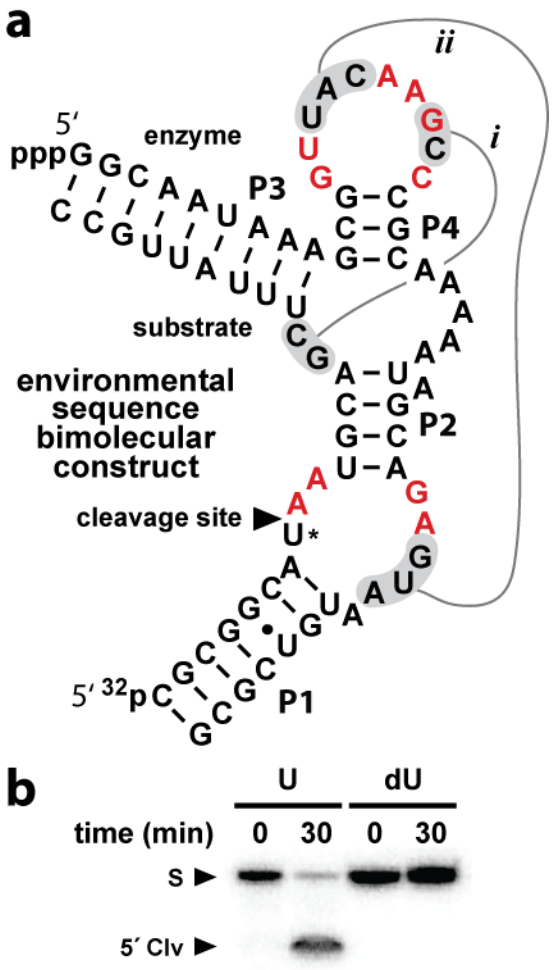
α = Arrange the 2′ oxygen, phosphorus, and 5′ oxygen in an in-line geometry

β = Neutralize the negative charge on the non-bridging phosphate oxygen

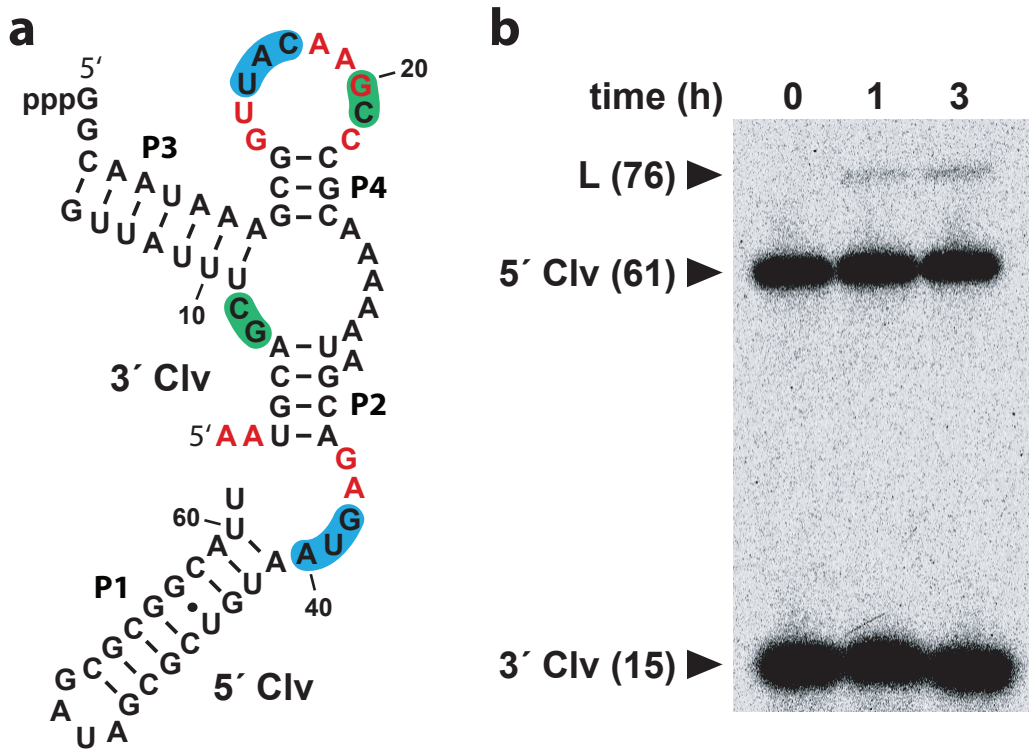
γ = Deprotonate the 2′ oxygen

δ = Neutralize the developing negative charge on the 5′ oxygen

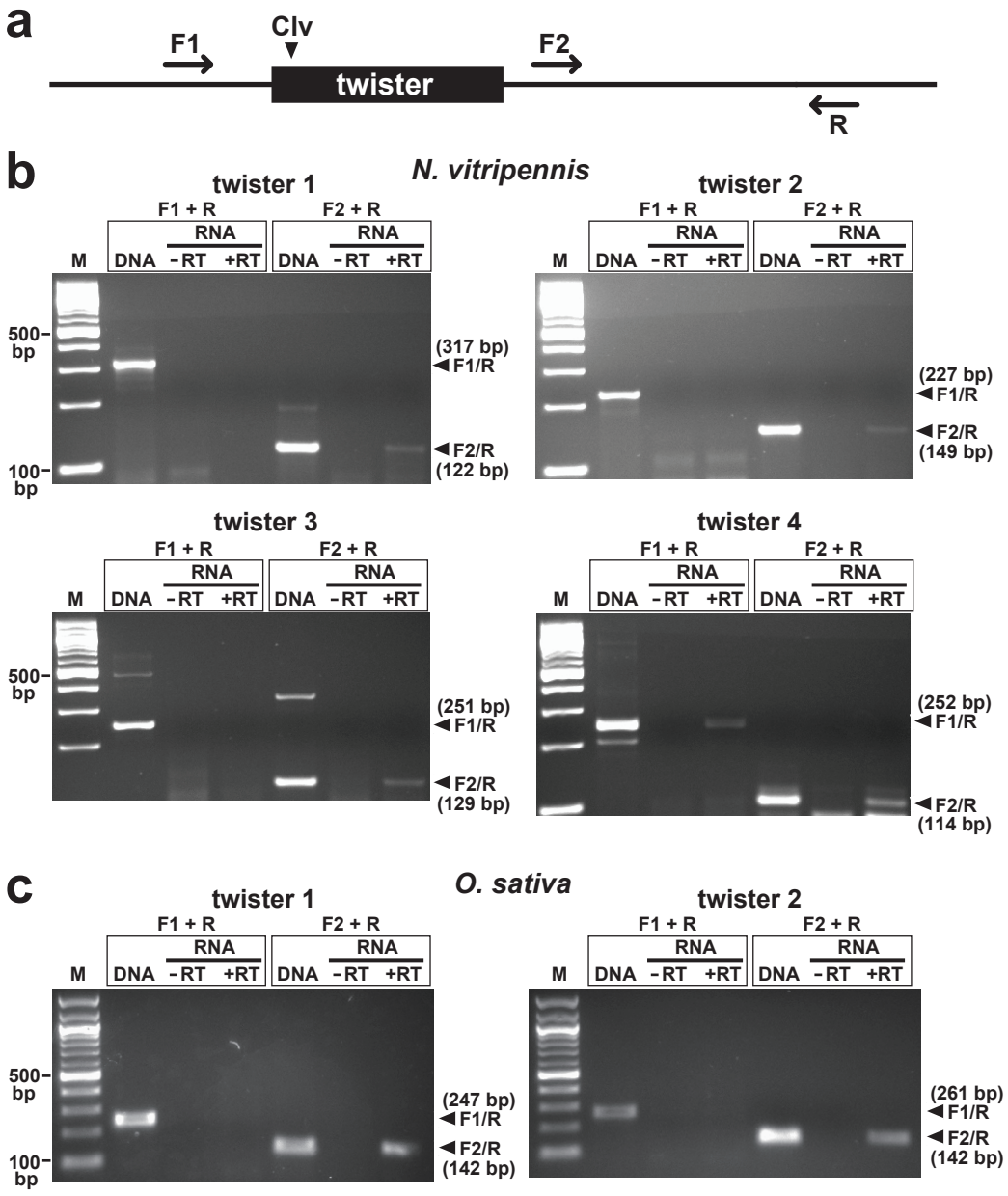
Supplementary Figure 4 | Internal phosphoester transfer mechanism for RNA cleavage. All six validated classes of natural self-cleaving ribozymes exploit this general cleavage mechanism. The end groups resulting from this manner of strand cleavage include a 2′,3′ cyclic phosphate on the 5′ cleavage fragment and a 5′ hydroxyl moiety on the 3′ cleavage fragment. Many natural and artificial RNA-cleaving enzymes, whether they are composed of RNA, DNA, or protein, employ multiple catalytic strategies in combination to accelerate this chemical transformation. The internucleotide linkage depicted corresponds to that of the twister ribozyme cleavage site consensus.

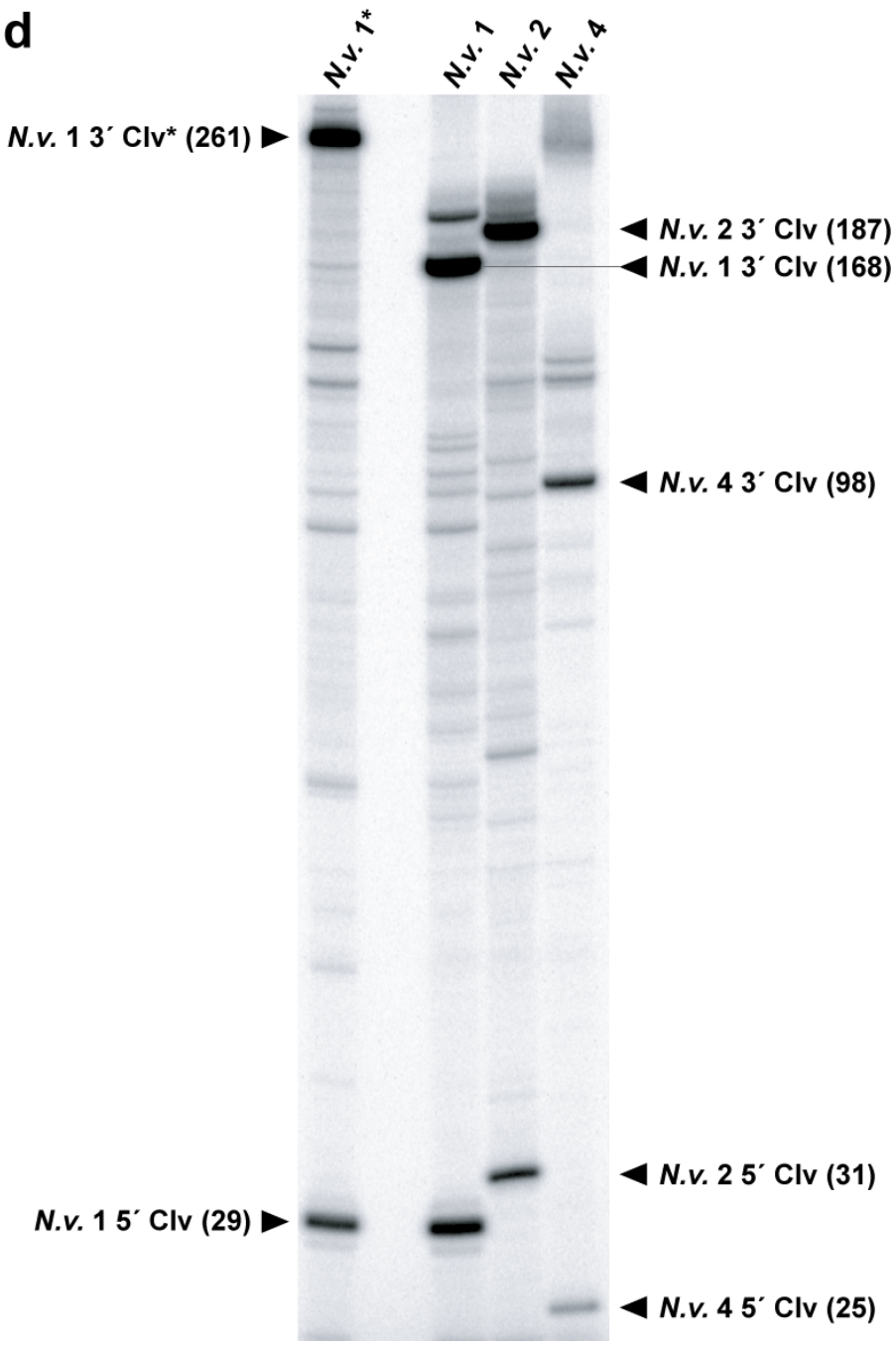


Supplementary Figure 5 | Twister ribozyme activity is not detected when the cleavage site 2′ OH group is absent. (a) Sequence and secondary structure of a bimolecular complex based on a type P3 twister ribozyme from environmental DNA (**Supplementary Fig. 1a**). The asterisk identifies the uridine located immediately 5′ to the labile phosphoester bond. The two guanosines at the 5′ end of the enzyme strand and the two cytidines at the 3′ end of the substrate strand are non-native. (b) Ribozyme cleavage assays using the bimolecular construct depicted in (a). U and dU refer to substrates that contain uridine and 2′-deoxyuridine, respectively, at the position in (a) indicated by the asterisk. Other annotations are as described in the legend to **Supplementary Fig. 2**.



Supplementary Figure 6 | Catalysis of ligation by a twister ribozyme. (a) Sequence and secondary structure of a complex consisting of individually purified 5′ and 3′ cleavage fragments (5′ and 3′ Clv), which were used as reactants in the ligation reaction. These fragments correspond to a type P3 twister ribozyme derived from environmental DNA (**Supplementary Fig. 1a**). Nucleotides in red are at least 97% conserved among identified twister ribozyme examples. The guanosine residues at the 5′ end of 5′ Clv are non-native and were added to facilitate transcription *in vitro*. Other annotations are as described in the legend to **Supplementary Fig. 1a**. (b) Ligation assay using internally ^{32}P-labeled fragments depicted in (a). Reactant and ligated product (L) RNAs were separated by denaturing 10% PAGE following incubations for the indicated times. Numbers in parentheses designate lengths in nucleotides. The small fractions of ligated products (0.4% and 0.7% after 1 h and 3 h, respectively) suggest that the internal equilibrium favors cleavage under the reaction conditions employed.


a
F1
Clv
twister
F2
R
b
N. vitripennis
twister 1
F1 + R
F2 + R
RNA
M
DNA
-RT
+RT
500 bp
100 bp
(317 bp)
F1/R
F2/R
(122 bp)
twister 2
(227 bp)
F1/R
F2/R
(149 bp)
twister 3
(251 bp)
F1/R
F2/R
(129 bp)
twister 4
(252 bp)
F1/R
F2/R
(114 bp)
c
O. sativa
twister 1
(247 bp)
F1/R
F2/R
(142 bp)
twister 2
(261 bp)
F1/R
F2/R
(142 bp)

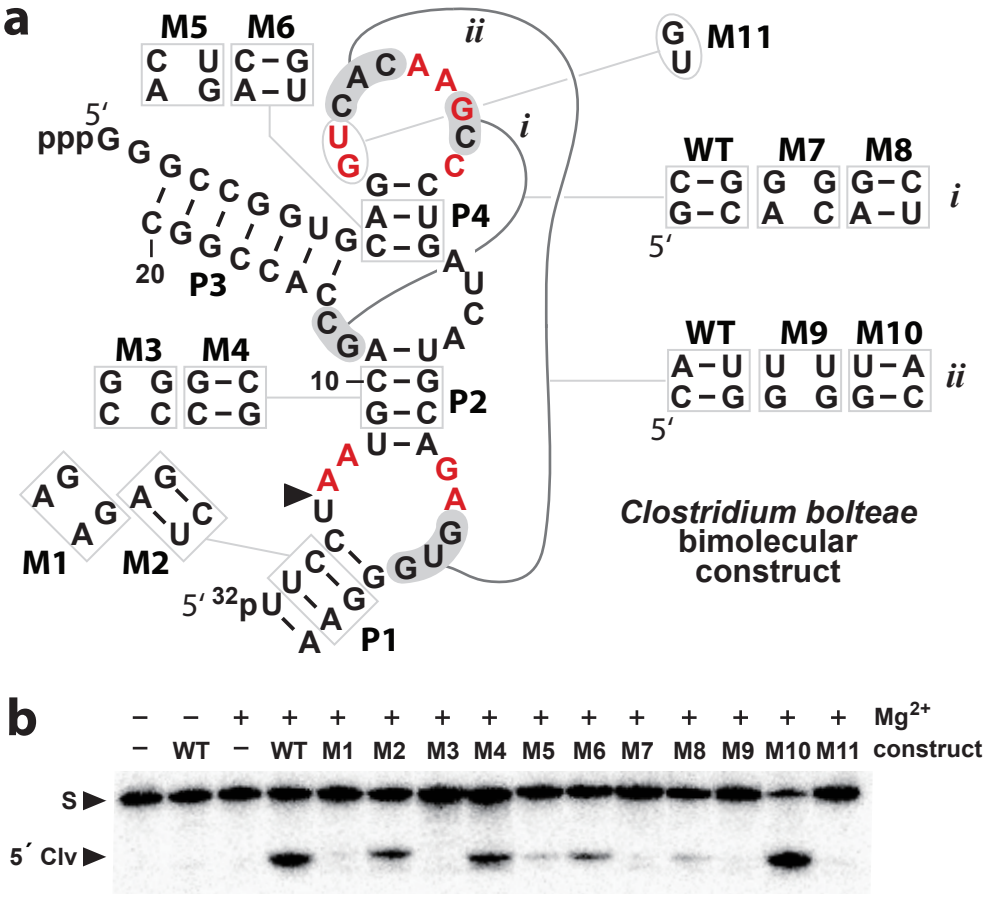

d
N.v. 1*
N.v. 1
N.v. 2
N.v. 4
N.v. 1 3´ Clv* (261)
N.v. 2 3´ Clv (187)
N.v. 1 3´ Clv (168)
N.v. 4 3´ Clv (98)
N.v. 2 5´ Clv (31)
N.v. 1 5´ Clv (29)
N.v. 4 5´ Clv (25)

e

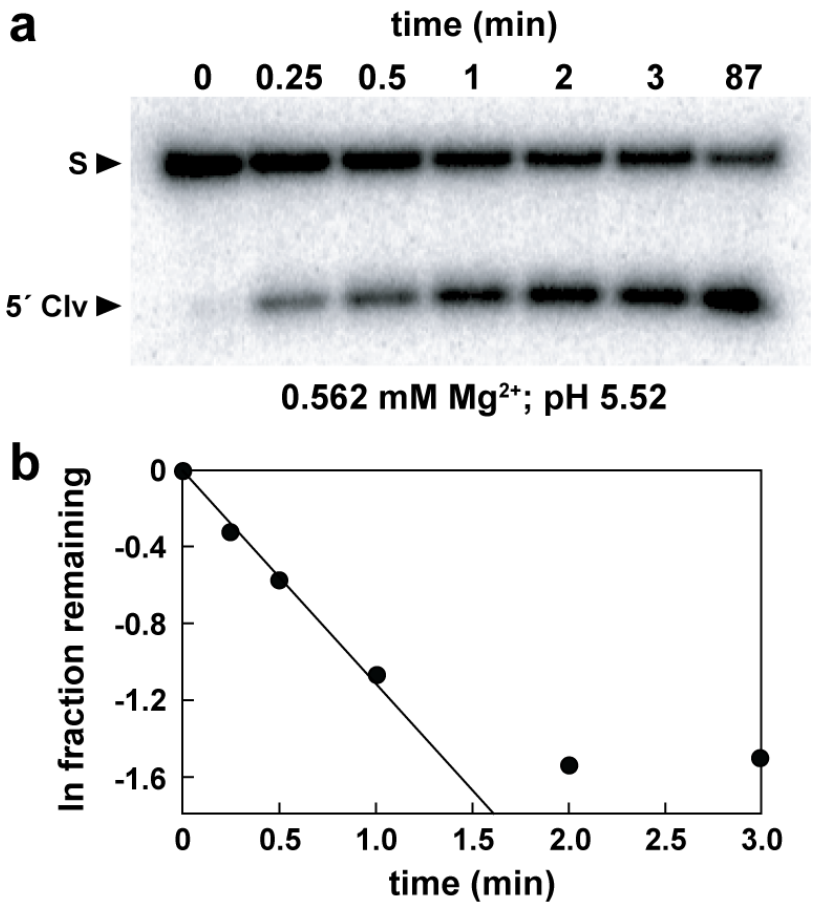
Twister ribozymes	Ribozyme coordinates	Locus-specific primers
N. vitripennis 1	NW_001820510.1/14932-15004	**F1:** CATAATATAAGGTTACTGTTGCTATACGGAAACCTG **F2:** TTGTGCGTTGAATAGATACTCGAGAACTGAG **F3:** TAATACGACTCACTATAGGAAACCTGAAAAATAGCAACACATTTTTAATGTGG **R:** GGTGAATCCTTAGACCTCGGTGGC **R1:** CAGTTCTCGAGTATCTATTCAACGCACAATTG
N. vitripennis 2	NW_001815459.1/1079284-1079348	**F1:** CTTGATACAAGTGTGAGATGTGCAAATGTTCC **F2:** CCCGATAAACGCAGAGAGCAAGTGC **F3:** TAATACGACTCACTATAGGTGTGAGATGTGCAAATGTTCCGCACTTG **R:** GTGATAATTACGCGTTTGTTATATAACACATGCATG
N. vitripennis 3	NW_001820349.1/671906-671853	**F1:** TACCTTGTCCCGTACTATTTCCCGTAC **F2:** CATAATCGCGCGCTTGAAATCGAACTATAAG **R:** CCACCTACTCGTTGAATATTCACCTTGTTATG
N. vitripennis 4	NW_001815555.1/7122-7175	**F1:** ATGATCTACCTTGTCCCGTACTATTTCCC **F2:** GCGCTTGAAATCGAACTATAAGTTCTAAAGACG **F3:** TAATACGACTCACTATAGGTACTATTTCCCGTACACACTTTTAATGTGGC **R:** CTGTTACGTAAATATACTGAAACGTCTGGTTTCC **R1:** CGCGATTATGATTTGTTAAATGAGGACGATTTTAC
O. sativa 1	NC_008404.1/1825405-1825354	**F1:** CTTGCTTCATATCTGTAGACGGAATG **F2:** CTGTGTCGGGGTGTTTTCGGTG **R:** GCATCGACTAATGAACTGCAAGCGC
O. sativa 2	NC_008405.1/1767469-1767418	**F1:** CAGTCAGAGGTGCTCCAATTGTTTTTTC **F2:** CCGGTTTCTTCAAGTTTTTGAGTTGTG **R:** CTGCTCTTTCTGAAGCTTATTGGGCG

Supplementary Figure 7 | Evidence for twister ribozyme function *in vivo*. (a) Relative positions and orientations of the primer sequences F1, F2, and R (forward 1, forward 2, and reverse, respectively) used to analyze twister RNAs occurring in *N. vitripennis* and *O. sativa* Japonica. The location of the cleavage site (Clv) is indicated. (b) RT-PCR analyses of twister RNAs from *N. vitripennis*. Total RNA was isolated from adult wasps and reverse transcribed using random hexamer primers. The resulting cDNA was PCR-amplified using F1/R or F2/R primer combinations (+RT). As negative controls, mock reactions that lacked reverse transcriptase served as templates for PCR amplification (–RT); positive controls employed wasp genomic DNA as template (DNA). Samples were analyzed by agarose gel electrophoresis. Arrowheads indicate the corresponding amplification products (with sizes in parentheses) generated for each of the twister representatives. A 100-bp ladder (M) was included as a size marker. (c) RT-PCR analyses of twister RNAs from *O. sativa* Japonica, as in (b). (d) The internally ^{32}P-labeled products generated during *in vitro* transcriptions of *N. vitripennis* twister ribozymes 1, 2, and 4 (*N.v.* 1, *N.v.* 2 and *N.v.* 4) from (b) were separated by denaturing 12.5% PAGE. The corresponding 5′ and 3′ cleavage products (5′ Clv and 3′ Clv, respectively) are

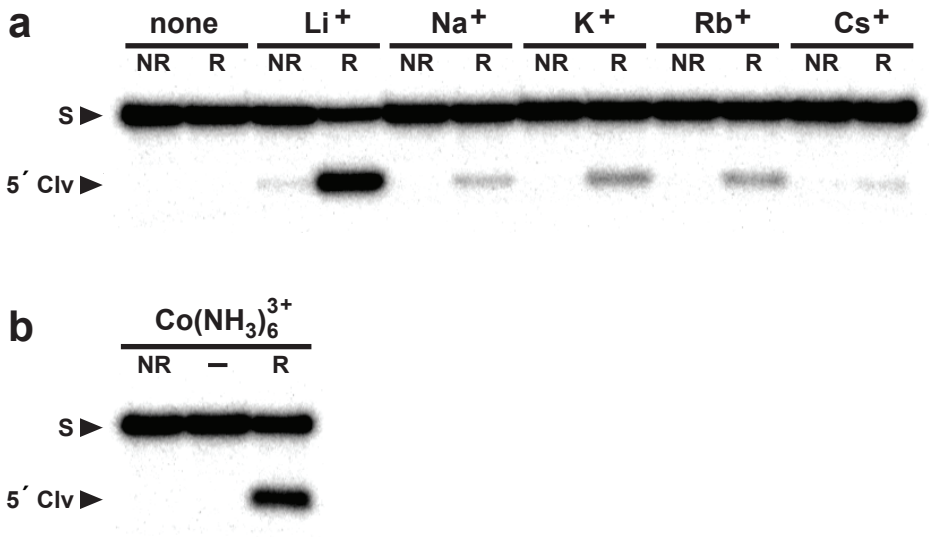
indicated, and the numbers in parentheses indicate fragment lengths in nucleotides. The *N.v.* 1* construct contains a 3′ end that is 93 nt longer than that of *N.v.* 1, but is otherwise identical. DNA templates for transcriptions *in vitro* were amplified using the following primer pairs: F3 and R (*N.v.* 1*), F3 and R1 (*N.v.* 1), F3 and R (*N.v.* 2), and F3 and R1 (*N.v.* 4). (e) Coordinates and primer sequences corresponding to the ribozymes that were subjected to RT-PCR analysis and to those that were transcribed *in vitro*. All oligonucleotide sequences are 5′ to 3′.



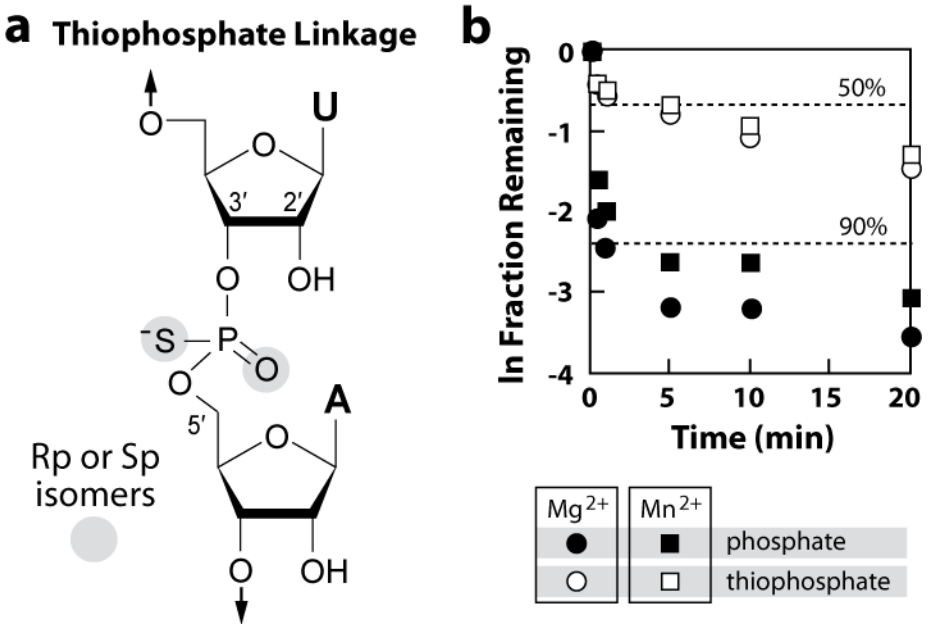
Supplementary Figure 8 | Experimental support for the secondary structure model. (a) Sequence and predicted secondary structure of an engineered bimolecular twister complex derived from *Clostridium bolteae*. The altered nucleotides in mutant constructs M1 through M11 are indicated. For each predicted pairing element, putative destabilizing and compensatory mutants were designed. Other annotations are as described in the legend to **Supplementary Fig. 2**. Non-native guanosine residues were encoded at the 5′ end of the enzyme strand to facilitate transcription *in vitro*. (b) Ribozyme cleavage assays for the constructs depicted in (a). Other details are as described in the legend to **Supplementary Fig. 2**, except that reactions were incubated for 1 min.



Supplementary Figure 9 | Representative time course and corresponding plot for a bimolecular twister ribozyme derived from an environmental sequence. (a) 5´ ^{32}P-labeled substrate RNA (S) and unlabeled enzyme RNA corresponding to a twister representative from environmental sequence (**Supplementary Fig. 5a**) were reacted as indicated under otherwise standard conditions. The 5´ cleavage product (5´ Clv) was separated from the uncleaved substrate by denaturing 20% PAGE. An extended time point (87 min) was included to correct for the fraction of substrate RNA remaining uncleaved after exhaustive incubation. (b) Corresponding plot of the natural logarithm of the corrected fraction remaining versus time. The negative slope of the line reflects the initial rate constant for the reaction without correction for the fraction of RNA remaining uncleaved after exhaustive incubation.



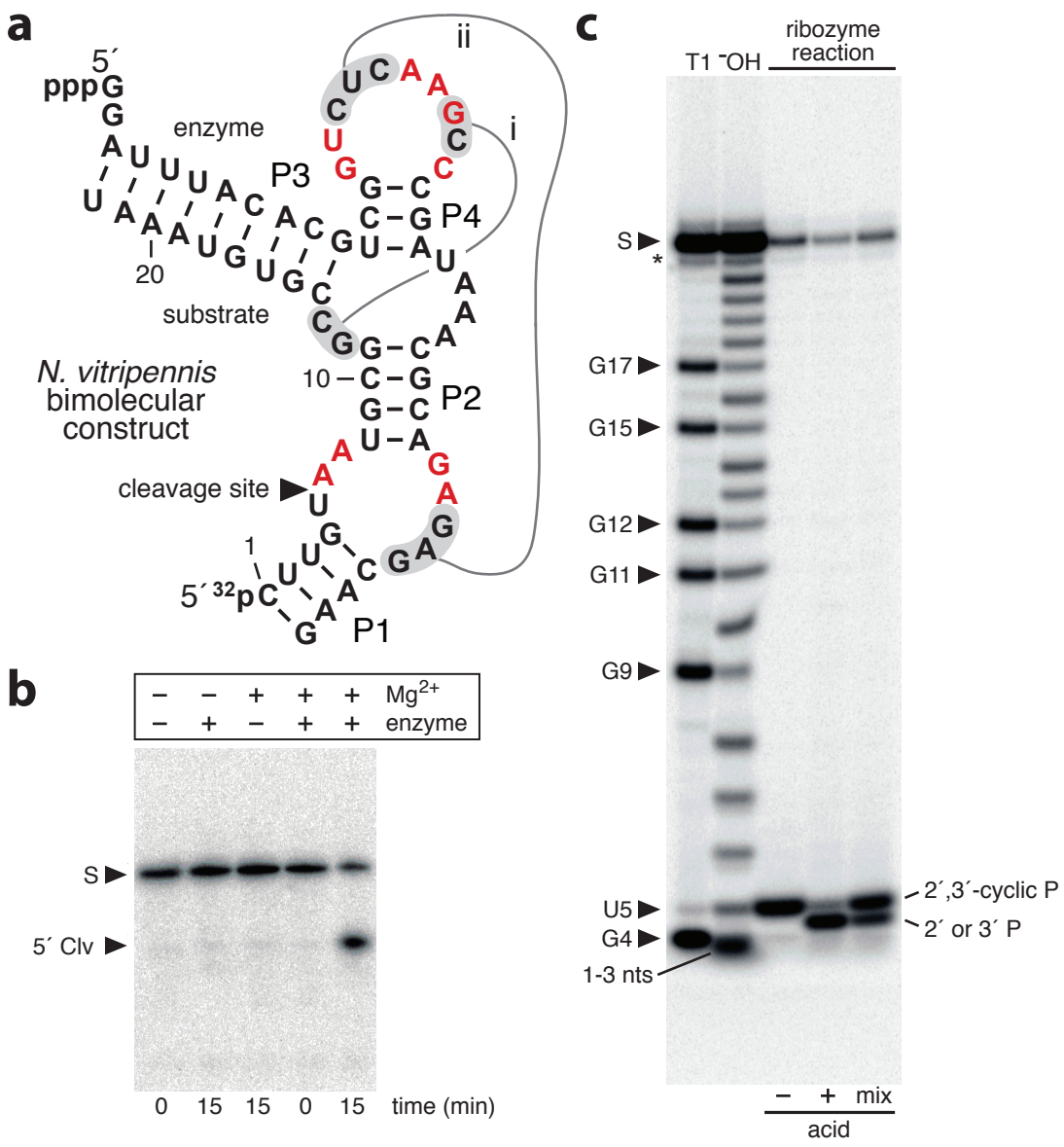
Supplementary Figure 10 | Divalent metal ions are not required for twister ribozyme activity. (a) An engineered bimolecular twister ribozyme complex derived from environmental sequence (**Supplementary Fig. 5a**) was reacted (R) for 1 min at room temperature in the absence (none) or presence of the indicated monovalent cations at concentrations of 1 M. To chelate contaminating divalent metal ions, 30 mM EDTA was also included in these reactions. For "no reaction" (NR) samples, five volumes of stop solution containing 90% formamide, 50 mM EDTA, 0.05% xylene cyanol and 0.05% bromophenol blue were mixed with the ribozyme complex prior to the addition of monovalent cations. Other details are as described in the legend to **Supplementary Fig. 2**. (b) Reactions of the same bimolecular complex were incubated in the absence (–) or presence of 1 mM cobalt hexammine chloride ($Co(NH_3)_6^{3+}$) at 23°C for 30 min. Other details are as described above.



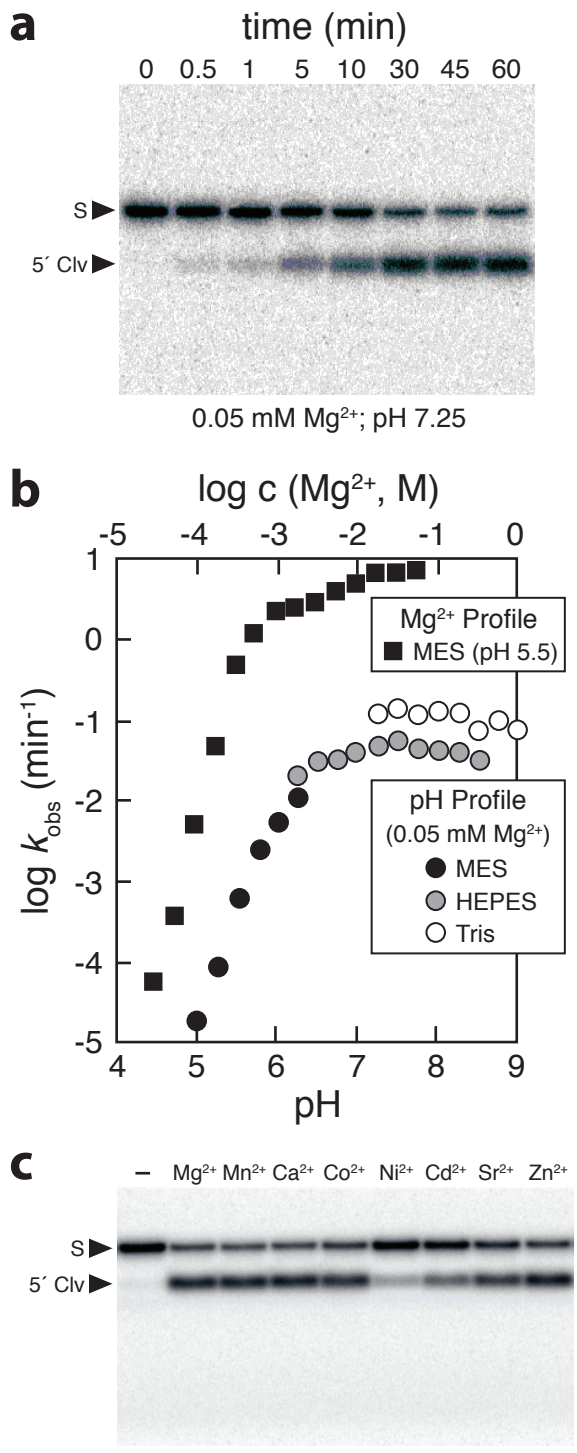
Supplementary Figure 11 | Twister ribozymes lack a metal-dependent thio effect. (a) Chemical structure of a phosphorothioate RNA linkage. Note that the substrate RNA bearing this modification is a mixture of two isomers (Rp and Sp; shaded atoms) at this position. (b) Plot of the natural logarithm of the fraction of substrate remaining uncleaved versus time using a bimolecular ribozyme complex derived from environmental sequence (**Supplementary Fig. 5a**). Cleavage assays were conducted in reaction mixtures containing either 1 mM $MgCl_2$ or 0.1 mM $MnCl_2$ under otherwise standard conditions. Dashed lines designate y-axis values that represent 50% or 90% processing of cleavable substrate. Data were corrected for the ~25% of substrate that remained uncleaved after exhaustive incubation.

In contrast to unmodified substrates, which are rapidly cleaved to ~90%, phosphorothioate-containing substrates undergo rapid processing only to ~50% before the k_{obs} decreases, suggesting that one of the two thiophosphate isomers is cleaved less efficiently than the other. However, for each substrate (unmodified or phosphorothioate-containing), the rate constants are nearly identical whether the reactions are performed in the presence of Mg^{2+} or more thiophilic Mn^{2+} cations, suggesting that the chemical mechanism does not involve direct metal ion coordination to cleavage site non-bridging phosphate oxygens. If inner-sphere interactions of this type were critical, Mn^{2+} cations would be expected to support faster rate constants under conditions where ribozyme chemistry is rate-limiting. Other self-cleaving ribozymes whose

reaction mechanisms appear to employ divalent metal ion coordination to non-bridging phosphate oxygens exhibit higher rate constants with phosphorothioate-containing substrates when Mn^{2+} ions are substituted for Mg^{2+} ions[34,57].



Supplementary Figure 12 | Sequence, structure and activity of a twister ribozyme from *N. vitripennis*. The data shown are identical to those in **Figure 3**, except that full-length versions of the gel images are included. The figure legends are also as in **Figure 3**.



Supplementary Figure 13 | Kinetic characteristics of a twister ribozyme derived from an environmental DNA sequence. The data shown are identical to those in **Figure 4**, except that full-length versions of the gel images are included. The figure legends are also as in **Figure 4**.